Volume 285, number 1, 25–27 FEBS 09863 July 1991

ADONIS 001457939100591N

Primary structure of myohemerythrin from the annelid *Nereis diversicolor*

Takashi Takagi[1] and Jos. A. Cox[2]

From the [1]Biological Institute, Faculty of Science, Tohoku University, Sendai, Japan 980 and *[2]Department of Biochemistry, University of Geneva, 1211 Geneva 4, Switzerland*

Received 9 April 1991

The metal-free form of *Nereis diversicolor* myohemerythrin was purified from whole animal extracts by trichloroacetic acid precipitation and ion exchange chromatography. The amino acid sequence of myohemerythrin has been determined. The protein is composed of 120 residues, possesses an unblocked N-terminus and is devoid of cysteine residues. It bears 62% sequence identity with *Themiste zostericola* myohemerythrin, the only other member of this subfamily sequenced to date. Within the family of hemerythrins, homology is particularly high in the segments involved in the binding of the two iron atoms and in the β-turn-rich N-terminal segment.

Myohemerythrin; Annelid; Primary structure

1. INTRODUCTION

Besides porphyrin-containing pigments such as the giant hemoglobins in the blood, annelids possess also oxygen binding proteins in which iron atoms are directly bound to the protein, namely extracellular hemerythrins in the vascular and coelemic system and intracellular myohemerythrins in muscle [1]. Whereas 4 hemerythrins are structurally well defined [2–5], the knowledge of the structure of myohemerythrins is restricted to that of the sipunculans *Themiste zostericola* [6–7]. The latter protein contains 118 residues and is mostly composed of a left twisted 4-α-helical bundle. Its active center contains 5 His, one Glu and one Asp, to which 2 iron atoms are directly bound. One oxide ion (O^{2-}) bridge and one oxygen complete the octahedral coordination of the non-heme iron atoms [8].

Our protocol for the purification of sarcoplasmic calcium-binding proteins from different invertebrates [9] revealed that impressive amounts of myohemerythrin can be isolated from annelids and purified in the metal-free form in a very short time. The purification and the complete amino acid sequence of the protein from *Nereis diversicolor* is reported here.

2. MATERIALS AND METHODS

2.1. *Enzymatic digestion and separation of peptides*

Myohemerythrin (10 nmol) was dissolved in 10 mM Tris-HCl buffer, pH 9.0, containing 4 M urea and digested with 2 μg of lysyl endopeptidase (Wake Pure Chemicals) at 37°C for 4 h or dissolved in 70% formic acid and cleaved with CNBr for 4 h at room temperature. The resulting peptides were purified by reverse-phase HPLC and their amino acid sequence determined from the N- to the C-termini. The overlaps of the C-terminal 3 lysylendopeptidase peptides were obtained by digesting myohemerythrin with *Staphylococcus aureus* V8 protease (Miles) in 100 μl of 50 mM phosphate buffer, pH 7.8 at 37°C for 4 h. Peptides were purified on a reverse-phase column with a linear gradient of acetonitrile.

2.2. *Amino acid analysis and sequence determination*

Peptides were hydrolyzed with gas of constant boiling HCl containing 0.2% phenol at 150°C for 1 h. Amino acid analysis was performed on a Hitachi L8500 amino acid analyzer with the *o*-phtalaldehyde method.

Amino acid sequences of protein and peptides were determined by using an automated sequencer (Applied BioSystem Model 477A on line with a Model 120A PTH-analyzer).

3. RESULTS AND DISCUSSION

3.1. *Isolation of myohemerythrin*

600 g of frozen whole animals were extracted twice with 2 vols of 20 mM Tris-HCl buffer, pH 7.5, 10 μM $CaCl_2$, 20 μM phenylmethanesulfonyl fluoride, 0.1 μg/ml pepstatin (buffer A). To the supernatant, trichloroacetic acid was added to a final concentration of 3%. The pellet was redissolved and the trichloroacetic acid step repeated once. The precipitate was dialyzed overnight against buffer A and put on a 5 × 20 cm column of DEAE-52 cellulose, equilibrated in the same buffer. Upon elution with a 2 × 1000 ml linear gradient of NaCl (0–0.3 M) myohemerythrin eluted at low conductivity as a double headed peak (Fig. 1), just before the peak of the sarcoplasmic calcium-binding protein [10], easily recognized by its concomitant Ca^{2+} containing peak. Fractions with conductivities between 2 and 4.5 millimho contained one single protein band upon urea- or sodium dodecyl sulfate-containing polyacrylamide electrophoresis. They were precipitated with 1% trichloroacetic acid, resuspended in buffer A and

Correspondence address: T. Takagi, Biological Institute, Faculty of Science, Tohoku University, Sendai, Japan 980

[1]The numerotation is based on the sequence of *Nereis* myohemerythrin as displayed in Fig. 3.

[2]The existence of major and minor variant forms has been reported [2].

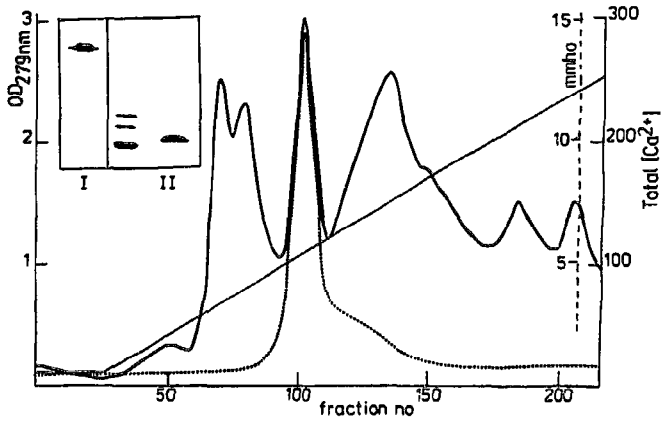

Fig. 1. DEAE-52 cellulose chromatography of trichloroacetic acid-treated extract of *Nereis*. Experimental conditions are reported in Section 3. Protein-bound Ca^{2+} (dashed line) was monitored by atomic absorption [9]. The thin line represents the ionic conductivity gradient in millimho. Inset: urea (I) and sodiumdodecyl sulfate (II) containing discontinuous polyacrylamide gel electrophoresis of pure myohemerythrin. Middle lane in II: molecular mass standards 16 949, 14 401 and 8149 Da.

rechromatographed on a 1.5 × 40 cm DEAE-52 cellulose column including an elution with a 2 × 250 ml gradient of NaCl (0–0.15 M). Myohemerythrin eluted as a single peak at a conductivity of 3 millimho and was electrophoretically pure (Fig. 1, inset). After extensive dialysis against water the protein was lyophilized. 600 g of whole *Nereis* yielded over 300 mg of pure protein.

3.2. *Characterization of myohemerythrin*

Both in sodium dodecyl sulfate-containing polyacrylamide electrophoresis (Fig. 1, inset) and in Sephadex G-75 gel filtration experiments (not shown) the protein behaves as a 13.5 kDa monomeric polypeptide, a property also reported for the *Sipunculans* myohemerythrin [1]. The absorption spectrum was flat from 300 to 600 nm, indicating that the protein was devoid of iron atoms, which was confirmed by atomic absorption spectrophotometry. It should be noted that after the first DEAE-52 cellulose chromatography the myohemerythrin-containing fractions were still brown; the second trichloroacetic acid treatment likely achieves complete removal of iron and causes elution in a single peak from the second DEAE-52 cellulose column. The metal-free protein is perfectly soluble after lyophilization and still shows Mg^{2+}-induced conformational changes when monitored by near UV difference spectroscopy.

2.3. *Amino acid sequence of myohemerythrin.*

The summary of data used for the sequence of *Nereis* myohemerythrin is shown in Fig. 2. The sequence of the N-terminal 61 residues was obtained by Edman sequencing on the intact protein and, when compaired with those stored in the Swiss Prot data bank, revealed the identity of the protein as a myohemerythrin (Dr A. Bairoch, personal communication). The majority of the total sequence was determined by using the cyanogen bromide and lysyl endopeptidase peptides. The overlaps of three lysyl endopeptidase peptides at the C-terminus needed also the use of V8 protease peptides.

In Fig. 3 the sequence of *Nereis* myohemerythrin was compared with that of *Themiste* as well as with the sequences of three hemerythrins. The two myohemerythrins display 62% sequence identity and 6% conservative replacements. *Nereis* myohemerythrin is two residues longer than the *Themiste* protein; two single-residue insertions were introduced in the β-turn segment

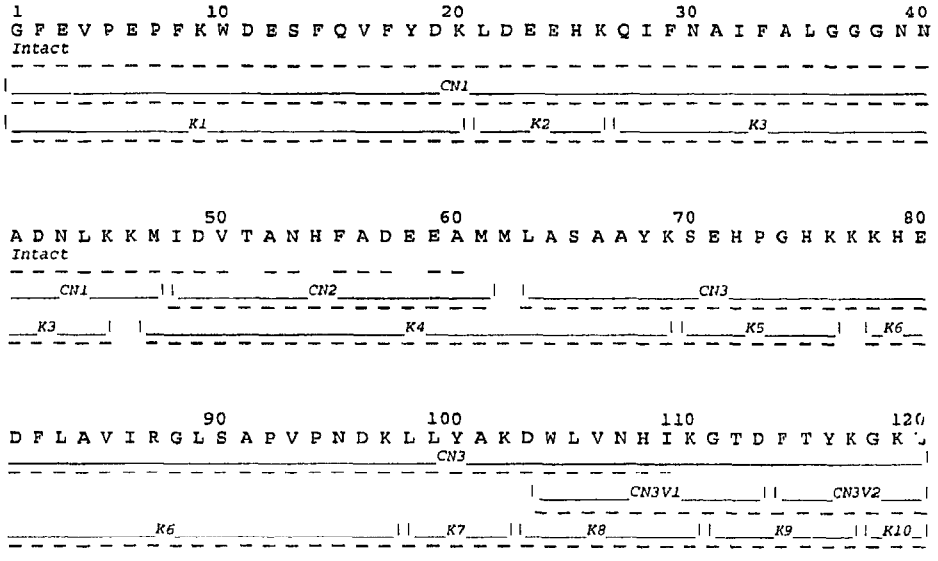

Fig. 2. Summary of data used to establish the amino acid sequence of *Nereis* myohemerythrin. Key: (- - -) automated sequencing; K, lysyl endopeptidase peptides; CN, cyanogen bromide peptides; V, *S. aureus* V8 protease peptides.

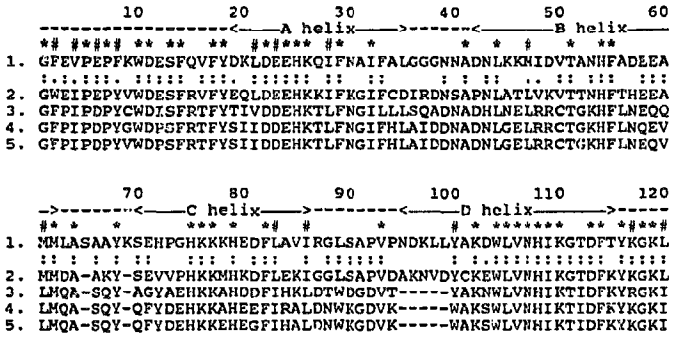

Fig. 3. Sequence similarities of myohemerythrins and hemerythrins. 1 and 2: myohemerythrins of *Nereis diversicolor* (this study) and *Themiste pyroides* [6]; 3, 4 and 5: hemerythrins of *Themiste dyscritum* [10], *Themiste zostericola* [3] and *Phascolopsis gouldii* [2][2], respectively. In myohemerythrins identical positions are represented by double dots, conservative replacements by single dots. * represents an identical residue in all the sequences, # a conservative replacement. The secondary structure assignment was taken from the known three-dimensional structure of *Themiste zostericola* myohemerythrin [8]. Dashes correspond to β-turn segments, ←helix→ to helical segments.

between the B and C helix in *Nereis* myohemerythrin. To improve the alignment with the hemerythrins a 5-residue gap was introduced in the beginning of the D helix in all the hemerythrins (note that the inserted segment in myohemerythrins is not conserved). Fig. 3 shows the particularly well conserved N-terminal arm (1 to 19) and the C-terminal segment. These segments are rich in β-turn. Also well conserved are the residues directly involved in the coordination of the iron atoms, namely H25, H54, H75, H80, H108, E58, D113 and their surroundings. Finally, most positions occupied by Trp, Tyr or Phe display extensive conservation. The fact that in the 5 sequenced members of the hemerythrins 42 positions are strictly conserved and 14 positions display conservative replacements indicates that strong evolutionary constraints have dominated in this family of proteins.

Acknowledgements: This work was supported by a Grant-in-aid for scientific research from the Ministry of Education and Culture of Japan and by the Swiss National Science Foundation Grant. We thank Isabelle Gerber for expert technical assistance and Dr Amos Bairoch for the computer analyses.

REFERENCES

[1] Klippenstein, G.L., Van Riper, D.E. and Oosterom, E.A. (1972) J. Biol. Chem. 18, 5959–5963.
[2] Klippenstein, G.L. (1972) Biochemistry 11, 372–379.
[3] Ferrell, R.E. and Kitto, G.B. (1971) Biochemistry 10, 2923–2929.
[4] Smith, J.L, Hendrickson, W.A. and Addison, A.W. (1983) Nature 303, 86–88.
[5] Stenkamp, R.E., Sieker, L.C. and Jensen, L.H. (1984) J. Amer. Chem. Soc. 106, 618–622.
[6] Klippenstein, G.L., Cote, J.L. and Ludlam, S.E. (1976) Biochemistry 15, 1128–1136.
[7] Sheriff, S., Hendrickson, W.A. and Smith, J.L. (1987) J. Mol. Biol. 197, 273–296.
[8] Hendrickson, W.A., Sung Co, M., Smith, J.L., Hodgson, K.O. and Klippenstein, G.L. (1982) Proc. Natl. Acad. Sci. USA 79, 6255–6259.
[9] Cox, J.A. and Stein, E.A. (1981) Biochemistry 20, 5430–5436.
[10] Loehr, J.S., Lammers, P.J., Brimhall, B. and Hermodson, M.A. (1978) J. Biol. Chem. 253, 5726–5731.